



Brain, Mind, World: Predictive coding, neo-Kantianism, and transcendental idealism

Zahavi, Dan

Published in:
Husserl Studies

DOI:
[10.1007/s10743-017-9218-z](https://doi.org/10.1007/s10743-017-9218-z)

Publication date:
2018

Document version
Peer reviewed version

Citation for published version (APA):
Zahavi, D. (2018). Brain, Mind, World: Predictive coding, neo-Kantianism, and transcendental idealism. *Husserl Studies*, 34(1), 47-61. <https://doi.org/10.1007/s10743-017-9218-z>

Forthcoming in *Husserl Studies*. Please quote from published version.

Brain, Mind, World: Predictive coding, neo-Kantianism, and transcendental idealism

Abstract:

Recently, a number of neuroscientists and philosophers have taken the so-called predictive coding approach to support a form of radical neuro-representationalism, according to which the content of our conscious experiences is a neural construct, a brain-generated simulation. There is remarkable similarity between this account and ideas found in and developed by German neo-Kantians in the mid-19th century. Some of the neo-Kantians eventually came to have doubts about the cogency and internal consistency of the representationalist framework they were operating within. In this paper, I will first argue that some of these concerns ought to be taken seriously by contemporary proponents of predictive coding. After having done so, I will turn to phenomenology. As we shall see, Husserl's endorsement of transcendental idealism was partially motivated by his rejection of representationalism and phenomenism and by his attempt to safeguard the objectivity of the world of experience. This confronts us with an intriguing question. Which position is best able to accommodate our natural inclination for realism: Contemporary neuro-representationalism or Husserl's transcendental idealism?

Can contemporary cognitive science and philosophy of mind profit from a closer engagement with the history of philosophy? There are many reasons why this question ought to be answered affirmatively. Fundamental reflections on and analyses of the mind-world relationship are not a new thing and it would be counterproductive to ignore resources found in the tradition. Doing so, might make one miss out on

important insights that in the best of circumstances would end up being rediscovered decades or centuries later. In the following contribution, I would like to exemplify this assessment by arguing that there are interesting similarities between the predictive coding framework that is currently in vogue in cognitive neuroscience and ideas found in and developed by German neo-Kantians in the mid-19th century. The point of this comparison will not simply be historical, however. As we shall see, some of the neo-Kantians eventually came to raise certain questions and harbour certain doubts about the cogency and internal consistency of the representationalist framework they were operating within. As I will argue, some of these concerns ought to be taken seriously by contemporary proponents of predictive coding. After having done so, I will turn to phenomenology. As we shall see, Husserl's endorsement of transcendental idealism was partially motivated by his rejection of representationalism and phenomenalism and by his attempt to safeguard the objectivity of the world of experience. This confronts us with an intriguing question. Which position is best able to accommodate our natural inclination for realism: Contemporary neuro-representationalism or Husserl's transcendental idealism?

1. Predictive coding

Recently, a number of neuroscientists and philosophers including Frith, Metzinger and Hohwy have defended different versions of the *predictive coding approach* (also sometimes referred to as *predictive processing theory* or *prediction error minimization theory*) and argued that it supports a form of radical neuro-representationalism; radical because the claim isn't simply that our access to the external world is mediated by neural representations, but rather that the world of experience is itself a representational construct. A central claim is that the brain doesn't process all the information it receives, but rather focuses its resources on unexpected input. To minimize costly surprises, however, the brain constantly

seeks to anticipate what signals its sensory organs will be receiving. To do this as efficiently as possible, the brain constructs internal models of the possible causes of those inputs. These models allow the brain to better predict likely inputs, and these predictions are then continuously compared with the actual incoming sensory inputs. In case of error, i.e., if there is a large discrepancy between the predicted and actual inputs, the model is revised and improved (Frith 2007: 126-127). On this view, the brain is seen as a hypothesis-testing organ. However, the only data the brain can work on are the internal effects of the stimulated sense organs. Their external causes remain hidden. Whatever evidence the brain has access to is consequently evidence available on the inside, so to speak. As Hohwy puts it, we can never crawl outside our brains and directly compare our representations and predictions with the external state of affairs (Hohwy 2016: 265). The content of our conscious experiences must therefore be considered a neural construct, a brain-generated simulation. As Frith puts it, "My Perception Is Not of the World, But of My Brain's Model of the World" (2007: 132). Whatever we see, hear, touch, smell, etc. is all contained in the brain, but projected outwards and externalized, such that we in normal life fail to recognize it as a construct and mistake it for reality itself (Metzinger 2009: 6-7). Colors often serve as the relatively innocent starting point:

The apricot-pink of the setting sun is not a property of the evening sky; it is a property of the internal *model* of the evening sky, a model created by your brain. The evening sky is colorless. The world is not inhabited by colored objects at all. [...] Out there, in front of your eyes, there is just an ocean of electromagnetic radiation, a wild and raging mixture of different wavelengths (Metzinger 2009: 20)

But what holds true of colors, also holds true for other familiar objects. The immediate objects of perception are in fact mental constructs. The visually appearing rose, the touched ice-cube, the heard melody, etc., are all brain generated representations, are all internal to and contained in the brain.

Given that we never have direct contact with external states of affairs – after all, the latter remains hidden behind the representational veil – we should reject all claims concerning the existence of a seamless tight coupling between mind and world. Hohwy speaks of the strict and absolute division between inner and outer and of the “evidentiary boundary” that secludes and separates the brain from everything beyond its boundary (Hohwy 2016). This even includes our own bodies, which are as hidden to us, as “very distal causes of sensory input such as the receding galaxies” (Hohwy 2016: 275).

Epistemically speaking, dualism reigns. But although the account on offer, as Hohwy admits, “entails skepticism” (Hohwy 2016: 265), although our representational filters “prevent us from seeing the world as it is” (Metzinger 2009: 9), this is not considered a serious cause of worry. As Metzinger reassures us, “an external world does exist, and knowledge and action do causally connect us to it” (Metzinger 2009: 23). Taking an instrumentalist line, Frith dismisses the concern that we can never know whether our internal model really matches with the external world. What is crucial is the possibility of prediction and control, i.e., that the model works. Whether it truthfully describes reality or not is irrelevant (Frith 2007: 136).

2. Neo-Kantianism

Intriguingly, one historical reference that keeps reoccurring in these recent writings is a reference to the work of Hermann von Helmholtz (e.g., Frith 2007: 41, 102; Clark 2013: 182; Hohwy 2013: 5; Hobson and Friston 2014: 8). A central reason for this is that Helmholtz is seen as an early defender of the view

that the task of perception is to infer the cause of sensation and that perception to that extent amounts to a form of hypothesis testing (Hohwy 2013: 77). Helmholtz was also a neo-Kantian, however, who argued that natural science had vindicated some of Kant's central insights (see De Kock 2014 & 2016). Helmholtz found an important source of inspiration for this claim in the work of the physiologist Johannes Müller, who had claimed that the properties of the external causes are not transmitted in a faithful and accurate manner to consciousness by our nerves. Indeed, so many intermediary steps and transformations occur on the way between the external cause and the experienced effect that any similarity or resemblance between the two can safely be ruled out. When we are cut by a knife, we do not feel an external object, but an internal pain. Likewise, when we see something, what we sense depends not only on the external causes, but also and far more significantly, on our own physiological constitution. As Müller wrote:

That which through the medium of our senses is actually perceived by the sensorium, is indeed merely a property or change of condition of our nerves; but the imagination and reason are ready to interpret the modifications in the state of the nerves produced by external influences as properties of the external bodies themselves (Müller 1842: 1059).

Helmholtz accepted this reasoning, and likewise argued that since the information about the external object is transformed beyond recognition on its way through the nervous system, what we end up perceiving is strictly speaking the internal effect rather than the external cause:

The result of [scientific] examination, as at present understood, is that the organs of sense do indeed give us information about external effects produced on them, but convey those

effects to our consciousness in a totally different form, so that the character of a sensuous perception depends not so much on the properties of the object perceived as on those of the organ by which we receive the information (Helmholtz 1995: 13).

I would interpret the sensation only as a sign of the object's effect. To the nature of a sign belongs only the property that for the same object the same sign will always be given. Moreover, no type of similarity is necessary between it and its object, just as little as that between the spoken word and the object that we designate thereby (Helmholtz 1995: 408).

In the end, Helmholtz took Müller's theory and the evidence he presented as a scientific confirmation of Kant's basic claim in *Kritik der reinen Vernunft* concerning the extent to which "we can have cognition of no object as a thing in itself, but only [...] as an appearance" (Kant 1998: B xxvi), and he argued that contemporary science on the basis of physiological evidence were reaching the same kind of insights as Kant had reached by a priori considerations. Our knowledge concerns reality as it is represented within ourselves, and not mind-independent reality as it is in itself, which remains unknowable. Although we are only ever presented with internal sensations, we can, however, infer that there must be an external world, since there has to be a "cause of our nerve excitation; for there can be no effect without a cause" (Helmholtz 1855: 41). The internal sensation is consequently referred to the external object as its hypothesized cause, through an (unconscious) process of inference to the best explanation.

Let us for a moment return to Kant. It has recently been claimed that Kant in his work on perception and cognition anticipated several core aspects of predictive coding theory, and that the latter's

insistence on the active and hypothesis driven character of perception might even be seen as an updated version of Kant's *Copernican revolution* (Swanson 2016: 1, 4).¹ Already early on, however, Kant's theory was subjected to criticism. One of the early highly influential objections to *Kritik der reinen Vernunft* was formulated by Jacobi in 1787. Jacobi argued that Kant's appeal and reference to the things in themselves were violating his own critical system. We can according to Kant, have no knowledge of the things in themselves, yet we are supposedly justified in asserting their existence. And although causality for Kant is a category of understanding, whose applicability is restricted to the realm of the phenomena, he still ascribes it to the things in themselves, which although they neither exist in space and time (which are pure forms of intuition), are still taken to affect us and thereby cause us to have the representations we do. As Jacobi then concludes,

however contrary to the spirit of the Kantian philosophy it may be to say that objects make impressions on the senses, and in this way produce representations, it is hard to see how the Kantian philosophy could find an entry point for itself without this presupposition, and make any kind of presentation of its doctrine. [...] [W]ithout this presupposition, I could not find my way into the system, whereas with it I could not stay there (Jacobi 2000: 173).

Jacobi's critical observation was shared by Fichte, who in an early 1796 text wrote with approval of Jacobi's complaint (Fichte 1988: 325). Fichte considered the notion of the thing in itself inimical to

¹ Swanson further argues that predictive coding theory can be "seen as a major step in the evolution of Kant's transcendental psychology" (2016: 10), though he also acknowledges that the former's evolutionary, computational and neuroscientific approach "goes beyond Kant's insights in ways that Kant could not have imagined" (2016: 11). One might wonder, though, whether the naturalism of predictive coding theory is not ultimately incompatible with Kant's transcendental framework. For a critical take on the possibility of naturalizing Kant, see Allison 1995. For a more general discussion of whether naturalism and transcendental philosophy can be reconciled, see Zahavi 2004, 2013.

Kant's system, as did other German idealists such as Hegel, who would go on to claim that the distinction between the thing in itself and the thing for us is a distinction we make, a distinction for us. As Hegel puts it in § 44 of *Wissenschaft der Logik*:

The *thing-in-itself* [...] expresses the object insofar as one *abstracts* from everything that it is for consciousness, i.e. from all determinations of sensation [*Gefühlsbestimmungen*] as well as from all determinate thoughts of it. It is easy to see what remains, namely the *complete abstractum*, something entirely *empty*, determined only as a *beyond* [...]. Equally simple, however, is the reflection that this *caput mortuum* is itself merely *the product* of thought, more specifically, [the product] of thought that has progressed to pure abstraction (Hegel 2010: 89).

Interestingly, we find a somewhat similar development regarding the status of the thing in itself among the physiological neo-Kantians. This is particularly evident in Friedrich Albert Lange's monumental and hugely influential *Geschichte des Materialismus und Kritik seiner Bedeutung in der Gegenwart* (1866). Lange initially endorsed Helmholtz' view that a scientific investigation of the physiology of the sense organs had provided partial confirmation of Kant's fundamental claims and could be interpreted as a corrected and improved form of Kantianism (Lange 1925: 3/202-203). The senses only provide us with access to the internally occurring effects of things, and not to the external things themselves. Accordingly, the world we experience must be seen as a product of our constitution. But, as Lange then pointed out, if we think this through it has some implications that are often overlooked. Firstly, the point just made obviously also holds true for our bodily organs. Our body, our sensory organs, our nerves, and even our

very brain all figure as elements in the world of experience and are consequently nothing but unfaithful images and signs of something unknown (Lange 1925: 3/219, 230). As Lange wrote,

The eye, with which we believe we see, is itself only a product of our ideas; and when we find that our visual images are produced by the structure of the eye, we must never forget that the eye too with its arrangements, the optic nerve with the brain and all the structures which we may yet discover there as causes of thought, are only ideas, which indeed form a self-coherent world, yet a world which points to something beyond itself (Lange 1925: 3/224).

Ultimately, however, it is not only our sensations, but also our conceptions that depend upon our physiological constitution and organization. That is, the internal mechanisms that give rise to our sensations are, according to Lange, equally responsible for producing our concept of physical matter and of an external reality (Lange 1925: 3/204). Just as a fish swims in the pond and cannot go beyond its limit, we live within the realm of our conceptions and representations. Even when speaking of the things in themselves, we are not moving beyond our own realm (Lange 1925: 3/226). The natural sciences have enjoyed an immense progress and there is no reason to doubt their accomplishments, but eventually one has to engage with the fundamental epistemological questions, and when doing so, Lange argued, one will come to the realization that the realism of the materialists is mistaken. The physiological investigation of our sense organs might seem to offer a thoroughly materialist account of knowledge acquisition, but in truth it undermines our belief in material, self-subsistent, objects, which is why materialism when thought through sufficiently radically reveals itself as a form of idealism (Lange 1925: 3/223).

In the end, Lange remained somewhat ambivalent about the thing in itself. Sometimes he seemed to claim that since appearances exist and since we need an explanation for their existence, we are entitled on the basis of an inference to the best explanation to posit things in themselves as the external causes of these appearances. Sometimes he argued that the valid application of the concept of cause is restricted to experience and that we are consequently not entitled to employ causal explanations beyond the limits of experience. And sometimes he held the view that while the concept of a thing in itself is intelligible, it is a concept that might well remain empty and that we will never know whether it actually has a referent (Edgar 2013: 107-108).

Whereas Lange somewhat ambiguously kept referring to the things themselves, Mach, who had been taught by students of Müller, took the final step a few years later. In publications from the 1880s, Mach also argued that a physiological investigation of the sense organs provided the foundations for a theory of knowledge, but on his view, it was safe to eliminate the things in themselves. The objects of experience are not (unfaithful) representations of a world beyond experience, rather all that exists are complexes of sensations (Mach 1895: 200-201), which is a view that in turn has been interpreted as either amounting to a form of neutral monism or as a robust form of phenomenalism.

3. The peculiar status of the brain

Let us with this overview in mind, return to the contemporary neuro-representationalists, who with some justification might be labelled neo-neo-Kantians (cf. Anderson and Chemero 2013: 204). One lesson to draw is that it can be hard to preserve the existence of an external world the moment one embarks on the representationalist journey. It is not altogether clear how Metzinger so confidently can declare that whereas the world of experience is a brain-generated illusion, the world as described by physics, the

world of electromagnetic radiation, is the world as it truly is. How are scientists able to transcend their internal world simulation? How do they manage to pull off this epistemic achievement? How can we ever know that our scientific theories, which are offered as explanations of the world of experience, really capture external reality? Why are they not merely elaborate cognitive extrapolations that remain as brain-generated and as internal to our constitution as everything else?

It might here be tempting to appeal to evolutionary considerations. Human beings inhabit a pre-existing natural world. To survive in that world, we have to have the capacity to acquire genuine knowledge of that world, since that is what enables us to undertake actions that can promote our survival. To put it differently, our cognitive machinery would never have evolved the way it did, would never have withstood selective pressure, unless it allowed our internal representations to match and track external reality. But this line of reasoning is hardly convincing. We cannot establish the trustworthiness of our representations by appealing to evolutionary theory, since the latter obviously presupposes the former.

Another strategy might be to point to the role of intersubjectivity. We do not develop our scientific theories individually, rather they are the result of a collaborative effort. They are constructed over time by the scientific community and can to that extent precisely be said to transcend our individual world simulation. The problem with this argument, however, should also be straightforward. Given the neurocentric disembodied framework, intersubjectivity is hardly something that can just be taken for granted. The very existence of other subjects is as much in doubt as the existence of external objects.

At this point, it might be argued that we should just face up to the fact that our cognition is limited, that absolutely secure knowledge of the external world is unattainable, that we throughout employ abductive reasoning, and that the findings of science remain our best bet. To argue in this way, however, is to miss the point. What is currently being questioned is not whether we are fallible cognizers,

or whether science is a worthwhile enterprise, but rather how the position in question can so confidently embrace scientific realism.

Ultimately, however, this is not the main worry. As we can learn from Lange, there is a much more troublesome problem, one that threatens to undermine the whole setup. Why are we in the first place considering the possibility that our objects of perception are internally generated constructs rather than real spatio-temporal objects? Because this is what our neuroscientific investigation of the brain suggests. But if this theory is to be taken seriously, it has to be consistent. It cannot merely be the visually appearing hammer, orange, and passport which are part of a brain-generated virtual reality, the same must hold true of the visually appearing brain, regardless of whether we perceive it ‘directly’ when performing brain surgery, or when looking at the colorful brain scans. After all, my empirical knowledge of the brain (and of neurophysiology) must be perceptually informed, and if we are to distrust the deliverance of the senses, then surely that must also hold for whatever they tell us about the brain. To put it in slightly different terms, it is hard to understand how one can motivate a general skepticism about perceptual experience on the basis of neuroscientific findings, since the latter – to some extent at least – presuppose the validity of the former. The main challenge, in short, is not how we can epistemically get out of the brain, but how we could possibly get into it in the first place. How do we at all know that there really is a brain? In order to enjoy any kind of initial plausibility, the neuro-representationalist account that we have been presented with must necessarily be half-baked. It asks us to abandon our naïve realism, our confidence in the objective existence of ordinary objects of experience, but it only does so half-heartedly. As Slaby & Heilinger rightly point out, since the whole theory is constructed around the workings of the brain, the model must presuppose that one worldly object is exempt from its skeptical concerns and that we can indeed observe and describe the brain as it really is. But

if indeed the brain as discovered by science is ‘real’ in the transcendent sense of the term, then it is hardly convincing that we stop there, claiming that of all we can see and perceive, *only* one single object, the brain, is ‘truly real’ and not just a representation, perceived as it is in itself (Slaby & Heilinger 2013: 89).

On the proposal in question, the world of experience is a representational construct generated by the brain. But this proposal is faced with an obvious dilemma. Either it also considers the brain a part of the world of experience, i.e., as a representational construct, and if so the account appears to be circular and explanatorily vacuous. After all, in explaining how representations come about, the account must appeal to a representation. The other option is to hold that the brain is not a representational construct. But not only is it unclear how the theory could be entitled to hold that view, it is also unclear why the brain should then be the sole exception.

If we really were cognitively confined in the way proposed by recent neuroscience, it is difficult to understand how the position in question could be coherently formulated, let alone justified. This is a very classical objection, one that quickly comes to mind to anybody familiar with the Kantian tradition. Perhaps the neuro-representationalists have a ready answer. To my knowledge, however, they haven’t yet confronted the challenge head-on.

It is important to distinguish between the neuroscientific models and the theoretical interpretations they are subjected to. Friston’s theoretical work (e.g., Friston 2010) constitutes an important source of inspiration for recent discussions of predictive coding and has been heralded as a “theory that is set to dominate the science of mind and brain in the years to come” (Hohwy 2016: 259). But whereas Hohwy has argued that the theory entails the existence of a strict boundary between cognitive systems and the environment and therefore rules out certain hypotheses concerning the mind’s

extended, embodied and enactive character (Hohwy 2016: 259), Clark has insisted that the theory is compatible with and supports a situated, embodied and distributed approach to cognition, and that the prediction error minimization theory introduces no worrisome barrier between mind and world (Clark 2013: 195, 198, Clark 2016). As Clark puts it, rather than saying that what we perceive is some internal representation, it is more correct to say that the brain's complex flow of sub-personal processing allows for a tight mind-world linkage and enables us to be perceptually open to the world itself (Clark 2013: 199). Clark, in turn, has subsequently been criticized for not going far enough in his criticism of representationalism, and recently proposals have been made that more directly seek to align Friston's core ideas with enactivism (Bruineberg et al. 2016, Gallagher & Allen 2016).² I will not here take a stand on what is the correct interpretation of Friston's work, nor on whether the predictive coding paradigm is necessarily committed to neuro-representationalism. But it should be obvious that the worries presented above are primarily directed against the representationalist construal and should be less troublesome for any non-representationalist alternative.³

4. Phenomenology

The naturalized epistemology presented to us by the neuro-representationalists might be driven by scientific findings, and to that extent appears to be on firm ground. But the former doesn't really do a

² One reason for these continuing disagreements is not only that people tend to employ notions such as 'direct', 'indirect', and 'immediate' in quite different ways, but also that there has been a tendency to conflate epistemic, experiential, and causal understanding of these terms. The presence of causal intermediaries does not necessarily entail that perception is experientially and/or epistemically indirect. Likewise, the fact that our perceptual access to spatio-temporal objects in the world is enabled and underpinned by various sub-personal mechanisms and non-conscious cognitive processes does not necessarily entail that we therefore fail to see the objects as they are in themselves. Rather, one might view the cognitive processing as that which makes it possible for us to experience those objects in the first place (Hopp 2011: 163).

³ Let me add that neuro-representationalism can obviously also be found outside of the predictive coding framework. For a prominent example, see Grush 2003.

very good job of securing the reality of the world. Most of us want to be realists. But it is not always that straightforward to determine what theory is best able to accommodate our realist intuitions. In fact, the neuro-representationalism just presented seems to fit rather neatly Kant's own definition of *empirical idealism*: "Idealism is the opinion that we immediately experience only our own existence, but can only infer that of outer things (which inference from effect to cause is in fact uncertain)" (Kant 2005: 294).

As Putnam has argued, it is the philosophers traditionally accused of idealism, namely the Kantians, the Pragmatists, and the Phenomenologists, rather than the metaphysical realists, who actually respect and honour our natural (inclination for) realism (Putnam 1987: 12). Kant famously held the view that his own *transcendental idealism* was not merely compatible with *empirical realism*, but that the latter required the former (cf. Allison 2015). Husserl defended the same view, though his understanding of transcendental idealism differed from Kant's in important ways, for instance, in its rejection of the thing in itself.

I will not here offer a full-scale interpretation of Husserl's transcendental idealism (but see Zahavi 2003, 2008, 2010, 2017), but it is worth pointing out that Husserl was familiar with physiological neo-Kantianism, and that his turn towards transcendental idealism was partially motivated by his rejection of both representationalism and phenomenalism, and by his efforts to safeguard the objectivity of the world of experience (Husserl 2003: xiii).

For Husserl, the world that can appear to us – be it in perception, in our daily concerns or in our scientific analyses – is the only real world. To claim that there in addition to this world exists a world-behind-the-scene, which transcends every appearance, and every experiential and theoretical evidence, and to identify this world with true reality is, for Husserl, an empty and countersensical proposition.

One articulation of these views can be found in *Ideen I*, where Husserl discusses and rejects the proposal that the perceptually appearing object is something merely subjective, a mere sign of or illusory depiction of its hidden cause. First of all, such a proposal fails to respect the categorical distinction between perceptual consciousness, on the one hand, and signitive and pictorial consciousness, on the other. Whereas we can think of objects, and thereby intend them, even in their absence, the situation is different when it comes to perceptual intentionality. Perceptual intentionality does not confront us with pictures or images of objects – except, of course, in so far as we are perceiving paintings or photographs – but with the objects themselves in their bodily presence. When we say that something *appears* perceptually, this should consequently not be understood in the sense that the perceptually given is a picture or sign of something else. In the case of signitive or pictorial intentionality, by contrast, we perceive something that then signifies or depicts something else. We are directed not at that which we perceive, but through it, at something else. Signitive and pictorial intentionality consequently presupposes perceptual intentionality and cannot explain it.

As Husserl points out in the lecture course *Ausgewählte phänomenologische Probleme* from 1915, nothing might seem more natural than to say that the objects I am aware of are outside my consciousness. When my experiences – be they perceptions or other kinds of intentional acts – present me with objects, one must ask how this could happen, and the answer seems straightforward: By means of some representational mediation. The objects of which I am conscious are outside my consciousness, but inside my consciousness, I find representations (pictures and signs) of these objects, and it is these internal objects that enable me to be conscious of the external ones. However, as Husserl then continues, such a theory is completely nonsensical. It conceives of consciousness as a box containing representations that resemble external objects, but it forgets to ask how we are supposed to know that the (mis)representations are in fact (mis)representations of external objects:

The ego is not a tiny man in a box that looks at the pictures and then occasionally leaves his box in order to compare the external objects with the internal ones etc. For such a picture observing ego, the picture would itself be something external; it would require its own matching internal picture, and so on *ad infinitum* (Husserl 2003: 106).

Representationalism notoriously courts scepticism: Why should awareness of one thing (an inner object) enable awareness of a quite different thing (an external object), and how can we ever know that what is internally accessible actually corresponds to something external? On Husserl's anti-representationalist view, however, the fit and link between mind and world – between perception and reality – isn't merely external or coincidental: "consciousness (mental process) and real being are anything but coordinate kinds of being, which dwell peaceably side by side and occasionally become 'related to' or 'connected with' one another" (Husserl 1982: 111). This claim is one that resounds throughout Husserl's oeuvre. As he, years later, would write in *Cartesianische Meditationen*, it is absurd to conceive of consciousness and true being as if they were merely externally related, when the truth is that they are essentially interdependent and united (Husserl 1960: 84). Husserl's idealism is not a reductive idealism. Husserl is not a phenomenalist that seeks to reduce the world to a complex of sensations. His opponent is not the dualist, but the objectivist, who claims that reality is absolute in the sense of being radically mind-independent. To deny the latter, to deny that the "universe of true being" lies "outside the universe of possible consciousness, possible knowledge, possible evidence" (Husserl 1960: 84), is not to say that reality literally exists in the mind, or that it is an intramental construction, but that reality is essentially manifestable, and therefore in principle available and accessible to consciousness. I think Sartre was

quite right, when he in a text from 1939 offered the following interpretation of Husserl's theory of intentionality:

Against the digestive philosophy of empirico-criticism, of neo-Kantianism, against all 'psychologism,' Husserl persistently affirmed that one cannot dissolve things in consciousness. You see this tree, to be sure. But you see it just where it is: at the side of the road, in the midst of the dust, alone and writhing in the heat, eight miles from the Mediterranean coast. It could not enter into your consciousness, for it is not of the same nature as consciousness. [...] But Husserl is not a realist: this tree on its bit of parched earth is not an absolute that would subsequently enter into communication with me. Consciousness and the world are given at one stroke: essentially external to consciousness, the world is nevertheless essentially relative to consciousness (Sartre 1970: 4).

On this interpretation, Husserl would be a natural ally of 4E approaches to cognition, rather than of the contemporary form of Cartesian internalism we see promoted by neuro-representationalists. When Varela, Thompson and Rosch in *The Embodied Mind* sought to negotiate a middle path between the Scylla of cognition as a recovery of a pregiven outer world (realism), and the Charybdis of cognition as the projection of a pregiven inner world (idealism), and when they argued that their stated goal was to move beyond the representationalist framework of inner vs. outer (1991: 172), they were articulating a form of enactivism that has many affinities with Husserl's position.

Thirdly, and in this context, most importantly, it is the perceived everyday object, and none other, that the "*physicist explores and scientifically determines following the method of physics*" (Husserl 1982: 119). It is the planetary bodies I observe in the sky, the water I drink, the flower I admire, etc. that

the natural scientist is also investigating and whose true nature, he seeks to determine in as exact and objective a manner as possible:

The physical thing which he [the physicist] observes, with which he experiments, which he continually sees, takes in his hand, puts on the scale or in the melting furnace: that physical thing, and no other, becomes the subject of the predicates ascribed in physics, such as weight, temperature, electrical resistance, and so forth (Husserl 1982: 120-121).

For Husserl, there are, in short, not two ontologically different objects, the appearing (intra-mental) object and the physical (extra-mental) object. Rather, there is only one appearing (extra-mental) object that carries categorically distinct but compatible sensuous and theoretical determinations. This is also why the findings of science and everyday experience, the scientific image and the manifest image, do not have to contradict each other. They can both be true according to their own standards. More generally speaking, the difference between the world of perception and world of science is not a difference between the world for us and the world in itself (falling in the province of phenomenology and science, respectively). It is a difference between two ways in which the world appears. The world of science is not an autonomous world, a world behind or below the manifest world. Rather the world that science studies is the same world as that of everyday experience, namely manifest reality, but now enriched and enlarged in scientific terms. This is also why it is phenomenologically absurd, as Heidegger once pointed out,

to speak of the phenomenon as if it were something behind which there would be something else of which it would be a phenomenon in the sense of the appearance which represents and expresses

[this something else]. A phenomenon is nothing behind which there would be something else. More accurately stated, one cannot ask for something behind the phenomenon at all, since what the phenomenon gives is precisely that something in itself (Heidegger 1985: 86).

For Husserl, physical nature makes itself known in what appears perceptually. The very idea of defining the really real reality as the unknown cause of our experience, and to suggest that the investigated object is a mere sign of a distinct hidden object whose real nature must remain unknown and which can never be apprehended according to its own determinations, is for Husserl nothing but a piece of mythologizing (Husserl 1982: 122). Rather than defining objective reality as what is there *in itself*, rather than distinguishing how things are *for us* from how they are *simpliciter* in order then to insist that the investigation of the latter is the truly important one, Husserl urges us to face up to the fact that our access to as well as the very nature of objectivity necessarily involves both subjectivity and intersubjectivity. Indeed, rather than being the antipode of objectivity, rather than constituting an obstacle and hindrance to scientific knowledge, (inter)subjectivity is for Husserl a necessary enabling condition. Husserl embraces a this-worldly conception of objectivity and reality and thereby dismisses the kind of skepticism that would argue that the way the world appears to us is compatible with the world really being completely different.

If Husserl's transcendental philosophical project today seems somewhat remote and removed from mainstream philosophy, this is primarily because of the massive onslaught of naturalism that we have witnessed in recent decades. What might be worth considering, however, is that some forms of naturalism, as exemplified by my discussion of neuro-representationalism, might be further removed from commonsense realism than Husserl's transcendental idealism.

5. Conclusion

Bennett and Hacker once pointed out that many of the neuroscientists who have castigated philosophy for its alleged failings – for not having accomplished anything scientifically worthwhile in its 2500-year history – remain unaware of the extent to which much of their own framework of thought has a philosophical heritage. Bennett and Hacker highlight the continuing influence of 17th-century epistemology and metaphysics (Bennett and Hacker 2003: 134, 407). As I have argued above, however, we shouldn't overlook the recent resurgence of ideas originating in the mid-19th century. In either case, however, the take home message is that it is unwise to ignore the tradition. We not only risk missing out on important insights, we also risk repeating old mistakes.⁴

REFERENCES

Anderson, M. L., Chemero, T. (2013), 'The problem with brain GUTs: conflation of different senses of "prediction"' threatens metaphysical disaster', *Behavioral and Brain Sciences* 36, 204–205.

Allison, H. E. (1995), 'On Naturalizing Kant's Transcendental Psychology', *Dialectica* 49 (2-4): 335-356.

⁴ I am indebted to helpful comments from Liesbet de Kock and John Michael.

- Allison, H., E. (2015), 'From Transcendental Realism to Transcendental Idealism: The Nature and Significance of Kant's 'Transcendental Turn'', in S. Gardner & M. Grist (eds.), *The Transcendental Turn* (Oxford: Oxford University Press), 20-34.
- Bennett, M.R., Hacker, P.M.S. (2003), *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.
- Bruineberg, J., Kiverstein, J., Rietveld, E. (2016), 'The Anticipating Brain Is Not a Scientist: The Free-Energy Principle from an Ecological-Enactive Perspective', *Synthese*, doi:10.1007/s11229-016-1239-1.
- Clark, A. (2013), 'Whatever Next? Predictive Brains, Situated Agents, and the Future of Cognitive Science', *Behavioral and Brain Sciences*, 36(3), 181–204.
- Clark, A. (2016), 'Busting out: Predictive Brains, Embodied Minds, and the Puzzle of the Evidentiary Veil', *Noûs*, doi:10.1111/nous.12140
- Edgar, S. (2013). The limits of experience and explanation: F.A. Lange and Ernst Mach on things in themselves. *British Journal for the History of Philosophy* 21/1, 100-121.
- Fichte, J.G. (1988), *Early Philosophical Writings*, trans. D. Breazeale. Ithaca: Cornell University Press.
- Friston, K. (2010), 'The free-energy principle: A unified brain theory?' *Nature Reviews. Neuroscience* 11(2): 127–138.

Frith, C. (2007), *Making up the Mind: How the Brain Creates Our Mental Worlds*. Oxford: Blackwell.

Gallagher, S., & Allen, M. (2016), 'Active inference, enactivism and the hermeneutics of social cognition', *Synthese*, doi.org/10.1007/s11229-016-1269-8

Grush, R. (2003), 'In defense of some 'Cartesian' assumptions concerning the brain and its operation.' *Biology and Philosophy* 18: 53-93

Hegel, G.W.F. (2010 [1830]), *Encyclopedia of the philosophical sciences in basic outline. Part I: Science of Logic*, trans. K. Brinkmann & D. O. Dahlstrom. Cambridge: Cambridge University Press.

Heidegger, M. (1985 [1925]), *History of the Concept of Time*, trans. T. Kisiel. Bloomington: Indiana University Press.

Helmholtz, H.v. (1855), *Das Sehen des Menschen*. Leipzig: Leopold Voss.

Helmholtz, H.v. (1995 [1853/1892]), *Science and Culture: Popular and Philosophical Lectures*. Chicago: Chicago University Press.

Hohwy, J. (2013), *The Predictive Mind*. Oxford: Oxford University Press.

Hohwy, J. (2016), 'The Self Evidencing Brain', *Noûs*, 50(2), 259–285.

Hobson, J.A. & Friston, K.J. (2014), 'Consciousness, dreams, and inference: The Cartesian theatre revisited,' *Journal of Consciousness Studies*, **21** (1–2), pp. 6–32.

Hopp, W. (2011), *Perception and Knowledge: A Phenomenological Account*. Cambridge: Cambridge University Press.

Husserl, E. (1960 [1929/1950]), *Cartesian Meditations: An Introduction to Phenomenology*, trans. Dorion Cairns. The Hague: Martinus Nijhoff.

Husserl, E. (1982 [1913]), *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First Book. General Introduction to a Pure Phenomenology*, trans. Fred Kersten. The Hague: Martinus Nijhoff.

Husserl, E. (2003), *Transzendentaler Idealismus. Texte aus dem Nachlass (1908-1921)*. Husserliana 36. Ed. R. Rollinger. Dordrecht: Kluwer Academic Publishers.

Jacobi, F.H. (2000 [1787]), 'On transcendental idealism', in B. Sassen (ed.), *Kant's early critics: The empiricist Critique of the Theoretical Philosophy* (Cambridge: Cambridge University Press), 169-175.

Kant, I. (2005 [1785-1789]), *Notes and Fragments*, trans. C. Bowman, P. Guyer, F. Rauscher. Cambridge: Cambridge University Press.

Kant, I. (1998 [1787]), *Critique of Pure Reason*, trans P. Guyer & A. W. Wood. Cambridge: Cambridge University Press.

Kock, L. D. (2014), 'Hermann von Helmholtz's Empirico-Transcendentalism Reconsidered: Construction and Constitution in Helmholtz's Psychology of the Object', *Science in Context*, 27(4), 709–744.

Kock, L. D. (2016), 'Helmholtz's Kant Revisited. The All-Pervasive Nature of Helmholtz's Struggle with Kant's Anschauung', *Studies in History and Philosophy of Science Part A*, 56, 20–32.

Lange, F.A. (1925 [1865]), *The History of Materialism and criticism of its present importance*, trans. E. C. Thomas. London: Kegan Paul.

Mach, E. (1895 [1882]), 'The Economical Nature of Physical Inquiry', in *Popular Scientific Lectures*, trans. by Thomas J. McCormack. Chicago: Open Court.

Metzinger, T. (2009). *The Ego Tunnel*. New York: Basic Books.

Müller, J. (1842), *Elements of Physiology II*, trans. W. Baly. London: Taylor and Walton

Putnam, H. (1987), *The Many Faces of Realism*. LaSalle, Illinois: Open Court.

Sartre, J.-P. (1970) [1939], 'Intentionality: A Fundamental Idea of Husserl's Phenomenology', *Journal of the British Society for Phenomenology*, 1(2), 4–5.

Slaby, J., Heilinger, J.-C. (2013), 'Lost in Phenospace: Questioning the Claims of Popular Neurophilosophy', *Metodo* 1/2, 83-100.

Swanson, L.R. (2016), 'The Predictive Processing Paradigm Has Roots in Kant', *Frontiers in Systems Neuroscience* 10:79, 1-13. doi: 10.3389/fnsys.2016.00079

Varela, F.J., Thompson, E., Rosch, E. (1991), *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: The MIT Press.

Zahavi, D. (2003), 'Phenomenology and metaphysics', in D. Zahavi, S. Heinämaa & H. Ruin (eds.): *Metaphysics, Facticity, Interpretation* (Dordrecht: Kluwer Academic Publishers), 3-22

Zahavi, D. (2004), 'Phenomenology and the project of naturalization'. *Phenomenology and the Cognitive Sciences* 3/4, 331-347.

Zahavi, D. (2008), 'Internalism, Externalism, and Transcendental Idealism', *Synthese* 160/3, 355-374

Zahavi, D. (2010), 'Husserl and the 'absolute'', in C. Ierna, H. Jacobs, F. Mattens (eds.): *Philosophy, Phenomenology, Sciences: Essays in Commemoration of Husserl* (Dordrecht: Springer), 71-92

Zahavi, D. (2013). 'Naturalized Phenomenology: A Desideratum or a Category Mistake?' *Royal Institute of Philosophy Supplement* 72, 23-42.

Zahavi, D. (2017), *Husserl's Legacy: Phenomenology, Metaphysics, and Transcendental Idealism*.
Oxford: Oxford University Press.